Routledge
Taylor & Francis Group

Journal of Military Ethics

ISSN: 1502-7570 (Print) 1502-7589 (Online) Journal homepage: www.tandfonline.com/journals/smil20

The AI Commander Problem: Ethical, Political, and Psychological Dilemmas of Human-Machine Interactions in AI-enabled Warfare

James Johnson

To cite this article: James Johnson (2022) The AI Commander Problem: Ethical, Political, and Psychological Dilemmas of Human-Machine Interactions in AI-enabled Warfare, Journal of Military Ethics, 21:3-4, 246-271, DOI: 10.1080/15027570.2023.2175887

To link to this article: https://doi.org/10.1080/15027570.2023.2175887

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 12 Feb 2023.

Submit your article to this journal

Article views: 13158

View related articles

View Crossmark data

Citing articles: 15 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=smil20

JOURNAL OF MILITARY ETHICS
2022, VOL. 21, NOS. 3–4, 246–271
https://doi.org/10.1080/15027570.2023.2175887

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

The AI Commander Problem: Ethical, Political, and Psychological Dilemmas of Human-Machine Interactions in AI-enabled Warfare

James Johnson

Department of Politics and International Relations, University of Aberdeen, Aberdeen, UK

ABSTRACT

Can AI solve the ethical, moral, and political dilemmas of warfare? How is artificial intelligence (AI)-enabled warfare changing the way we think about the ethical-political dilemmas and practice of war? This article explores the key elements of the ethical, moral, and political dilemmas of human-machine interactions in modern digitized warfare. It provides a counterpoint to the argument that AI "rational" efficiency can simultaneously offer a viable solution to human psychological and biological fallibility in combat while retaining "meaningful" human control over the war machine. This Panglossian assumption neglects the psychological features of human-machine interactions, the pace at which future AI-enabled conflict will be fought, and the complex and chaotic nature of modern war. The article expounds key psychological insights of human-machine interactions to elucidate how AI shapes our capacity to think about future warfare's political and ethical dilemmas. It argues that through the psychological process of human-machine integration, AI will not merely force-multiply existing advanced weaponry but will become *de facto* strategic actors in warfare – the "AI commander problem."

KEYWORDS

Political psychology; artificial intelligence; human-machine interaction; military ethics; autonomous weapons

Introduction

This article explores the key features of the ethical, moral, and political dilemmas associated with human-machine socio-technical interactions in artificial intelligence (AI)-enabled warfare.[1] It draws insights from cognitive psychology, political philosophy, and scientific-technological approaches to consider the confluence of military ethics, emerging technology, and human psychology. Can AI solve the ethical, moral, and political dilemmas of warfare? How might AI-enabled warfare affect our thinking about the ethical-political dilemmas and practice of war?

I argue in this article that through the psychological process of human-machine integration, AI (especially machine learning (ML)) will not merely force-multiply existing advanced weaponry but will likely become *de facto* strategic actors (planners,

CONTACT James Johnson james.johnson@abdn.ac.uk Department of Politics and International Relations, University of Aberdeen, King's College Aberdeen, Aberdeen, AB24 3FX, United Kingdom

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

warfighters, tacticians) in warfare – the "AI commander problem."[2] The diminished role of human commanders in the trajectory, controllability, and consequences of war is ontologically, ethically, and morally problematic. As scientific-technological innovations such as AI are developed as a panacea to the Clausewitzian fog friction of "human" war, and to solve the ethical-political dilemmas of warfare, it is easy to lose sight of the underlining social, political, ethical, and psychological contexts of war (Paret 1985).

Much of the present debate has revolved around ethical and legal concerns about fielding lethal autonomous robots (or "killer robots") into armed conflict. The literature contains both pessimism and optimism about the trajectory of human-machine interaction in war; even in technological-scientific circles, deep-seated worries about control and bias are widespread (Russell 2019; Sauer 2021; Gill 2019; Schmitt 2013). Less attention, however, focuses on the ethical (or "techno-ethical"), moral, and psychological dilemmas associated with the intersection of technology and warfare (Schwarz 2018a; Clark 2019; Dobos 2020; Bousquet 2018; Emery 2022; Roff 2014). The article fills a critical gap in discussing complex socio-technical interactions between AI and warfare. In doing so, it provides a valuable counterpoint to the argument that AI 'rational' efficiency can simultaneously offer a viable solution to humans' psychological and biological fallibility in combat while retaining "meaningful human control" over the war machine (Arkin 2009; Hagerott 2014).[3] The article also argues that framing the narrative in terms of "killer robots," and similar tropes, misconstrues both the nature of AI-enabled warfare and its ability to replicate and thus replace human moral judgment and decision-making.

How has AI altered our understanding of war (and ourselves)? The article offers three psychological insights into human-machine interactions to illuminate how AI will influence our capacity to think about the political and ethical dilemmas of contemporary warfare: (1) Human biological and psychological fallibility and the dehumanization of AI-enabled war; taking humans psychologically further away from the act of killing; (2) human psychology and cognitive bias in human-machine interaction and use of military force; and (3) the emergence of a techno-ethics of war in AI-enabled warfare and the implications for moral responsibility of war.

These insights address why, how, and to what effect human-machine interactions may harbinger *de facto* AI commanders in future warfare. Some of these insights relating to future war are necessarily speculative; only by extrapolating present trends in AI-enabling technology can we elucidate the potential implications of the current trajectory to draw logical (or illogical) conclusions. While we cannot escape our time, we can use theories and empirical analyses as tools to serve us in critical inquiry.

The article is organized into three sections. The first section frames the argument by contextualizing the broader dovetailing of humanity and technology with human-machine interactions. It considers why and how humans become so entangled with the machines and the emergent complex socio-technical system, the roots of military techno-ethics, and the notion of riskless, frictionless war. It describes AI technology as a new manifestation of this socio-technical trend. It argues that outsourcing human consciences in war-making – in an illusionary bid to solve the ethical, moral, and political dilemmas of war – risks eroding the vital link between humanity and war. This section also engages with the various counterarguments that challenge the view that

replacing humans with machines is *necessarily* a bad idea (the "AI optimists"). Humans, for instance, make mistakes, often act irrationally, and are predisposed to atavistic instincts including violence, immorality, and dehumanization (Haslam 2006; Brough 2007).

The second section considers the psychological features of human-machine interaction. Specifically, it unpacks several human biases – the illusion of control, heuristic shortcuts (the Einstellung effect, existence bias), and automation bias – which can make commanders prone to misuse or overuse military force for unjust causes. It also discusses the potential effects of these biases in the broader political drive for a technological *deus ex machina* for predictability and centralized control over warfare.

Finally, the third section considers the potential implications of pursuing the means of perfecting riskless and frictionless war with technologies like AI for military ethics and moral responsibility in war. It contextualizes the debates surrounding the encoding of human ethics into machines with AI technology. It also examines the role of human emotion, which gives us a sense of reason and deliberation, influences our decisions, and shapes our responses to ethical and moral dilemmas – situations in which no desirable outcome is obvious. Can human ethics be programmed into algorithms? And if so, how might humans retain their ethics and values if moral responsibility is outsourced to AI?

Dehumanization of AI-enabled war: machines hollowing out humanity?

Henry David Thoreau famously observed that combatants' service to the nation is "not as men mainly, but as machines, with their bodies" (Thoreau 1906, 359). An idea fast gaining prominence is that humans will soon become the Achilles heel in the emergent AI-enabled techno-war regime and will be inexorably distanced from the battlefield and eliminated by "rational," efficient, and autonomous weapons systems (Schwarz 2018a, 156). Intelligent machines will soon no longer need humans acting as autonomous agents. Instead, war-making (goal-setting, on-board-targeting, ROEs, mission command, completion success reporting, etc.) is increasingly being outsourced (whether consciously or inadvertently) to the judgments and predictive insights of algorithms. The logical end of this slippery slope is – or could at least be – a *de facto* AI commander, whereby the act of killing, and thus the responsibility attached to agency, is outsourced to machines. The emergent complex socio-technical system – to make war faster, more lethal, asymmetric, and efficient – is being accomplished through a fundamentally psychological process of human-machine integration.

Why have humans become so entangled with machines? The history of human-machine interactions and techno-ethics can be understood as a manifestation of the broader evolutionary dovetailing of humanity and technology. A key feature of human evolutionary history has been the pursuit to artificially augment our physical and mental capabilities (e.g. human vision using ground glass, hydrate silicon, and fiber optics). One could also call this the sacrificing of our physical strength for intellectual upgrades. Recent neurological studies demonstrate that the human brain adapts throughout life – and not just during childhood as previously thought – to fully incorporate new technology to exploit its potential. This trajectory lends credibility to cognitive philosopher Andy Clark's notion of a human technology symbiosis to explain how our sense of

self (or "human agency") is determined in part by our relationship with technology (Clark 2003).

The roots of military techno-ethics, which has enabled war at a distance, and the politically seductive notion of riskless war, can be traced back to the Western Enlightenment assumptions of visualization and empiricism.[4] Specifically, with the osmosis of rationalization of vision and mathematization of space maps to assumptions relating to verifiable facts and empirical data, "truth" became only what the eyes verified as "reality" (Clark 2003, 144). Because of the moral and political imperatives of the use of force and the exigencies of war, the pursuit of riskless war is both reasonable and commendable. In the context of technologically enhanced weapons, particularly when used in asymmetric war conditions, this drive threatens the basic foundations of Just War upon which the moral justifications for violence and killing in war precariously rest (Renic 2019). Christopher Coker writes: "Our ethical imagination is still failing to catch up with the fast-expanding realm of our ethical responsibilities" (Coker 2008, 152). This expanding realm is set to widen further as technology gains in autonomy, lethality, and "intelligence."

The increasingly complex entwining of technology and mankind, with its "multiplying [of] human strength" and "inhumanity and destructive effectiveness," challenge the assumptions that underpin ethics and morals in war (Arendt 1970, 53). At the bleeding edge of this human-machine fusion, AI technology augurs a new manifestation of an omniscient technological solution to the ethical-political dilemmas of war. On the larger question of the ethics of warfare in modernity, Harvard historian Drew Gilpin Faust argues that the "*seductiveness of war* derives in part, from its *location on this boundary of the human, the inhuman, and the superhuman*. Its fascination lies in its ability at once to allure and repel, in the paradox that thrives at its heart" (Howard 2011, emphasis added). The danger with the development of technology such as AI is the possibility that humans may seek to reconcile this paradox by unwittingly outsourcing our consciences of using lethal force to non-human agents who are ill-equipped to fill this void. According to Duncan MacIntosh, "by taking decisions to kill away from the soldier we will save his conscience by "moral off-loading" … maybe we can offload the moral burden onto autonomous weapons systems that will do the killing for us, thereby sparing us unbearable guilt" (MacIntosh 2015).

Shannon French and Anthony Jack conceptualize two categories of "dehumanization" in war. The first is "animalistic dehumanization" (or "sub-humanization"), which characterizes the enemy as inferior and creates psychological distance by generating contempt, disgust, or hatred. The second is "mechanistic dehumanization" (or "objectification"), which, by contrast, equates the enemy with inanimate objects (resulting in expressions such as, e.g. "neutralize the target"). "Mechanistic dehumanization" generates cold indifference, which, like "animalistic dehumanization," creates the psychological distance that permits combatants to kill without hesitation or compunction (French and Jack 2015, 165–195). French and Jack's "mechanistic dehumanization" is instructive for this exploration.

If technologies like AI draw combatants further away from the battlefield (both physically and psychologically), they risk becoming conditioned to view the enemy as inanimate objects "neither base nor evil, but also things devoid of inherent worth" (Brough 2007, 160–161). If the "emotional disengagement" associated with a mechanistically

dehumanized enemy is considered conducive for combat efficiency and tactical decision-making, the production of controlled and banal socio-technical interactions devoted to moral emotions (remorse, guilt, shame, compassion, etc.) will be seen as ethically and morally lamentable (French and Jack 2015, 186).[5] As Hannah Arendt warned: "the development of *robot soldiers* … would *eliminate the human factor completely* and, conceivably, permit one man with a push button to destroy whomever he pleases" (Arendt 1970, 50, emphasis added).

In collaboration with the International Human Rights Clinic at Harvard Law School, the Human Rights Watch 2012 report *Losing Humanity* argued that "emotionless robots could serve as tools of repressive dictators seeking to crack down on their own people without fear their troops would turn on them" (Human Rights Watch 2012). However, while human emotions can restrain humans, they can also unleash the basest instincts of humanity, including those emotions associated with "animalistic dehumanization." Contemporary history is replete with tragic case studies of unchecked human emotions causing human suffering, such as Rwanda, the Balkans, Darfur, Afghanistan, and most recently, Ukraine (Lerner and Keltner 2001, 146–159). Therefore, the argument against the deployment of AI and autonomous weapons centered on the absence of human emotions oversimplifies both the complexities of human emotion and cognitive states and the psychologically nuanced nature of human-machine interactions, and is empirically flawed. Instead, elucidation is needed on the psychological and ethical impact of the dehumanization of war to address the danger of humanity being hollow-out (spatially, temporally, and corporeally) by intelligent machines (Coker 2013, xviii).

How might AI-enabled war upend Thucydides's inseparable union of war and human nature (Thucydides 1996)? Some argue that framing humans as biologically and psychologically fallible (see below) and inferior to intelligent machines further removes humans from the moral, ethical, and legal decision to use lethal force (Emery, 179–197). In this new AI-enabled techno-military regime, humans will become further intertwined with machines, "not merely to be better but to meet the *quasi-moral mandate of becoming a rational and progressive product*: ever-better, ever-faster, ever-smarter, *superseding the limited human corporeality, and eventually the human*" (Schwarz 2018a, 196, emphasis added). This argument is underpinned by an assumption that human (biological and psychological) fallibility can be enhanced and ultimately supplanted by AI systems, making war more rational, predictable, and controllable and, in turn, eroding the inseparability of human nature and war envisioned by Thucydides.[6] Although the notion of AI superseding and replacing humans in war is highly speculative and contested, the psychological, ethical, and moral implications of increasing AI-enabled human-machine interactions are ripe for empirical and theoretical exploration.

Human psychology and cognitive bias in human-machine interaction

What psychological insights can be garnered from human-machine entanglement? Several human psychological factors associated with human-machine interaction can cause commanders to misconstrue unjust conflicts (i.e. conflicts contrary to the just war tenets of *jus ad bellum* and *jus in bello*) as legitimate and place undue confidence in dispassionate machine rationality, thereby making militaries more predisposed to

exaggerate their ability to control events and more prone to resort to military force. This section considers three cognitive biases and psychological predispositions that might compel commanders to misuse or overuse military force for unjust causes: the illusion of control, the quest for order and predictability, heuristic shortcuts, automation bias, and AI and techno-rationalization.

The illusion of control

People's belief in their ability to control pure chance events is a recurrent finding in experimental psychology (O'Creevy et al. 2003, 53–68). According to behavioral studies, decision-makers in competitive, adversarial, violent contexts, where decision-makers are most emotionally attached to a particular outcome – war-making being an obvious example – are more prone to overstate their ability to control of events. The studies reveal that individuals engaged in "skill-oriented" situations (competition, choice, interaction, etc.) are prone to exaggerate their control and deny or misjudge the existence of chance, contingency (the "just world hypothesis"), and luck (Thompson 1999, 187; Dobos 2020, 91). Moreover, and most pertinent to human-machine interactions, the studies found that if participants were familiar with a simulated task, such as a wargaming exercise, they were even more prone to the "illusion of control" (Dobos 2020, 311). That is, people in a "skills-orientation" situation or process tend to conflate chance and contingency with a skill (or "skills cues") they have developed. The illusory promise of control and certainty is dispelled neither by enhanced expertise, know-how, or sober reflection; instead, these measures often have the reverse effect (Kahneman and Renshon 2009, 79–96).

It is easy to imagine how this might play out in the context of AI-enabled warfare. For instance, the leaders of state A contemplate using AI-enhanced cyberattacks (e.g. data poisoning, malware, or denial of service) against state B's command and control networks in response to B's unlawful annexation of sovereign territory along a shared border. Intelligence officers, flanked by geospatial technology, AI-enabled autonomous drone swarms on intelligence, reconnaissance, and surveillance (ISR) missions, and AI "big-data" analytics warn leaders of the risk of inadvertent escalation and wider damage – for example, damaging state B's nuclear retaliatory capabilities or civilian infrastructure. Recent wargaming simulations conducted by state A that run similar contingencies to those unfolding in the real world increases the leader's confidence (i.e. "skill cues") in their ability to control events on their terms. This is known as "escalation dominance" (Kahn 1965). In the real world, however, once the order has been given to deploy offensive cyber weapons – and without the means of recall, reliably signal, or proportion responsibility in cyberspace – the capacity of state A to control escalation and limit broader damage is limited (Acton 2020). Given these dangers and challenges, how are humans expected to retain meaningful control over the command-and-control decision-making process?

Quest for order and predictability

The history of command in warfare can be understood as a continuous quest for order over chaos and complexity (or "chaoplexic warfare") and to impose control and

predictability amid uncertainty (Bousquet 2008). In battle, the side best able to understand the various contingent elements that comprise the strategic environment in which war is fought – for example, battlefield awareness, the adversary's intentions, and the activities of one's allies and one's own forces – have invariably prevailed (Creveld 2003, 264). Military historian John Keegan (1976, 18–19) notes that the central purpose of military training "is to reduce the conduct of war to a set of rules and a system of procedures – and therefore to make orderly and rational what is essentially chaotic and instinctive." Against the backdrop of the re-emergence of strategic great power competition, AI technology has become the newest currency for commanders to reinvigorate their scientific quest to impose predictability and certainty on the modern battlefield (Lieber 2000; Jervis 1978; Johnson 2021).

The political momentum behind the drive for complete predictability and centralized control over warfare was catalyzed by the threat of nuclear Armageddon during the Cold War and manifested in the computer revolution and the greater use of analytical tools, sensors, radar technology, and data processing systems (Talmadge 2019; Bousquet 2008; Wiener 1967). US General William Westmoreland in 1969 encapsulated a vision of a future scientific way of warfare: "On the battlefield of the future, enemy forces will be located, tracked, and targeted almost instantaneously through the use of data links, computer-assisted intelligence evaluation, and automated fire control" (Westmoreland 1969, 215–223). Although humans have broken with Cartesianism – that scientific knowledge can be derived *a priori* from "innate ideas" via deductive reasoning – we are still seduced by the allure of science and controlling war in many ways. AI-enabled weapons are, therefore, symptomatic of a cumulative longer-term effort by militaries to use technology to tame chance and irradicate uncertainty in chaoplexic warfare (Boulanin 2020).

Westmoreland's vision inspired a new generation of techno-military concepts and approaches, including cybernetic warfare, network-centric warfare (NWC), and the broader concept of a revolution in military technology, to create a centralized, frictionless, and automated warfare (Cohen 1996; Cebrowski 1999; Alberts and Hayes 2003). It is noteworthy that most military technological "revolutions" have been justified on similar grounds, that is, morally justifying the pursuit of a specific technology (e.g. nuclear and chemical weapons) to make war more efficient and less brutal, even when those technologies ultimately have a detrimental humanitarian impact (Caron 2020). In fact, most predictions have gone wrong when they have overestimated the technological factor – and underestimated the human one (O'Hanlon 2018). Moreover, technological advances can also place a heavy burden on existing ethical norms, legal regulations, practices, and notions such as proportionality, responsibility, and meaningful human control.[7] Shifts in the threat environment can cause our moral vocabulary to adapt in lockstep with the capabilities and functions of machines we invent.

Driven by the assumption that machines will obviate fallible, emotional, and irrational human combatants, the enchantment of an AI-enabled micromanaged battlefield – combining AI-augmented ISR, autonomous weapons, and real-time situational awareness – has renewed the military adoption of Westmoreland's techno-military regime as a panacea to the Clausewitzian fog and friction of war in the digital age (Arquilla and Ronfeldt 1997; Kissinger, Schmidt, and Huttenlocher 2021; US ONR 2014). According to Robert Castel, this endeavor reflects "a grandiose technocratic rationalizing dream of

absolute control of the accidental understood as the irruption of the unpredictable. In the name of this *myth of absolute eradication of risk*" they may inadvertently manufacture novel risks (Castel 1991, 288, emphasis added). Frequent accidents involving human-machine interactions demonstrate that a human-in-the-loop is not a panacea, particularly when it is challenging to distinguish civilian objects from combatants and military objectives (Aircraft Accident Investigation Board Report 1994).[8]

Commanding war in complex and uncertain strategic environments entails more than voluminous, cheap (and often biased) data sets, and inductive machine logic. Until AI systems can produce testable hypotheses or reason by analogy and deductively reason (using "top-down" logic) like humans, they will not understand the real world and not be fully able to make decisions in non-linear, complex, and uncertain environments (Norvig 2014).[9] Commanders' intentions, the rules of law and engagement (e.g. the principle of proportionality), and the exhibiting of ethical and moral leadership in the execution of strategic objectives are critical features of ethical, moral, and tactically effective military decision-making (e.g. highly context-dependent targeting decisions) (Roff 2014, 211–227). If we hold AI-ML systems to be incapable of properly performing these intrinsically human traits, the role of human agents in "mission command" – the implicit communication and bond of trust between tactical leaders and the political-strategic leadership – will be even more critical in future AI-enabled warfare (Goldfarb and Lindsay 2022; Kramer 2015; Beyerchen 1992–1993).

Heuristic shortcuts

Arguably, as geopolitical and technological-deterministic forces spur militaries to embrace AI – in the pursuit of fleeting first-mover advantages of speed, lethality, and scale – commanders' intuition, emotion, and latitude will be needed more than ever before to cope with the unintended consequences, organizational friction, strategic surprise, and dashed expectations associated with the implementation and assimilation of military innovation (Horowitz 2010). This problem may be compounded by a cognitive propensity of individuals and organizations to "fixate on one particular kind of solution to a problem due to one's exposure to, or familiarity with, that solution" – that is, a heuristic shortcut to solve problems as efficiently as possible, known as "the Einstellung effect." A related concept is "Maslow's hammer," encapsulated by Abraham Kaplan's analogy: "Give a small boy a hammer, and he will find that everything he encounters needs pounding" (Dobos 2020, 88–89).

In a military context, this cognitive bias can make decision-makers prone to use capabilities just by virtue of possessing them, having invested time, energy, or political capital and resources in their acquisition. In his report to the UN Human Rights Council of lethal autonomous robotics (LARs), legal scholar Christof Heyns notes:

> Official statements from Governments with the ability to produce LARs indicate that their use during armed conflict or elsewhere is not currently envisioned … subsequent experience shows that *when technology that provides a perceived advantage over an adversary is available, initial intentions are often cast aside*. (Heyns 2013, emphasis added).

This bias can impair the ability of decision-makers to make impartial and objective risk assessments of the likelihood of operational success, thereby, unbeknownst to the

decision-maker, distorting the appraisal of competing alternatives to problems. As John Kleinig notes in the case of police militarization: "Equipment purchased 'just in case' … suddenly finds a use in situations that do not readily justify it" (Kleinig 2015). While it would be an exaggeration to claim that the Einstellung effect will distort every judgment on the use of military force, it may nonetheless contribute to the generation of false positives about the necessity for war (Dobos 2020, 90).

An overdue focus on AI-enabled speed and tactical efficacy is particularly perturbing in crisis-management situations (e.g. nuclear brinkmanship), where inadvertent escalation risks loom large, and humans rely on conceptual, analytical, and conscious deliberation (or "System 2' thinking) to make fast and reflexive judgments in stressful situations for cognitive closure (or "System 1' thinking) (Kahneman 2011). For example, the US DoD's 2022 Joint All-Domain Command and Control (JADC2) strategy report proposes integrating AI-ML technology into command and control (C2) systems to speed up the "decision cycle" relative to adversary abilities (US Department of Defense 2022). The JADC2 report obfuscates the possible strategic implications of automating the decision-making cycle for tactical gains and AI's illusionary clarity of certainty (Johnson 2022). AI-ML algorithms are unable to effectively mimic "System 2" thinking (or "top-down" reasoning) to make inferences from experience and abstract reasoning, with perception being driven by cognition expectations, which is a critical element in safety-critical contexts where uncertainty and imperfect information require adaptions to novel situations (Bahdanau, Cho, and Bengio 2014; Bengio, Lecun, and Hinton 2021).

AI and techno-rationalization redux

State-of-the-art AI language-based machine learning systems in production today, such as OpenAI's GP3 and DeepMind's Gato, have been successfully used in context-specific reasoning tasks such as summarizing documents, generating music, classifying objects in images, and analyzing protein sequences. However, these systems are limited by the amount of information they can "remember" while executing a given task – or the problem of "continued learning" (Wiggers 2021).[10] As a result, whether writing an essay or controlling a robot or automobile, these systems often fail to recall what they have learned from a training data-set; systems must be constantly reminded of the knowledge they have gained or risk becoming "stuck" with their most recent "memories" derived from their training data.

Several high-profile public displays of AI systems in gaming and simulated virtual environments highlight the unpredictability and inexplicability of techno-rationalization (AlphaStar Team 2019; Pawlyk 2020). One of the most worrisome features of AI-enabled warfare is a reductionist scientific view of war that may delude commanders into thinking that war can be controlled and predicted by objective, neutral, and rational machines. During stressful and fast-moving crises (e.g. anti-access area-denial contested zones), using AI decisions to provide an aura of objective legitimacy in place of prudence may result in the opportunistic misuse of machine logic to validate legally or ethnically questionable behavior or justify existing practices rather than justifying existing practices by seeking out alternatives. Predictive policing studies in the United States, for example, have demonstrated how a combination of biased AI training data-sets and an

overreliance on machines made officers prone to dismiss or not seek out contradictory information in preference of algorithmically generated judgments which they choose to accept as fact (Millar 2014; Meijer and Wessels 2019).

Some just war scholars argue that the military has already ceded some authority over *jus in bello* to machines that provide technical analysis to ensure decision-makers stay within the law, which has further defused the moral responsibility for actions from senior military leaders and military lawyers to "computer-assisted expertise" (Crawford 2013, 233). The authority and agency ceded to intelligent machines – based on an assumption of technical superiority, practical utility, and neutrality in decision-making – will likely be compounded by developments in AI-enhanced capabilities (e.g. LAWS, ISR, big-data analytics, robotics, and cyberweapons) that further entwine humans with machines, making potential mishaps (either human or machine) go unseen and unethical decisions more difficult to detect and thus contest (Schwarz 2018a, 159). In sum, advanced weapons systems augmented by AI technology challenge existing notions of agency and contestability, potentially innovating novel and increasingly autonomous ways to kill in a new techno-military regime.

As AI-enabled capabilities become assimilated into military doctrine, operational concepts, and strategic culture, militaries will be prone to unreflectively assign positive moral attributes to the resultant AI-enabled techno-military regime. Recent empirical psychology studies corroborate David Hume's hypothesis that people have an immediate favorable response to what is already established; they tend to "imbue the status quo with an unearned quality of goodness, *in the absence of deliberative thought*, actual experience or reason to do so" (Eidelman et al. 2009, 765, emphasis added; Hume 1992). In other words, if a technology such as AI exists, people will assume that "what is, ought to be." Its existence is unquestioned and justified and becomes the cognitive default. This is what we know as "existence bias" (Eidelman et al. 2009).

Furthermore, the blind pursuit of AI-enabled tactical efficiency also risks downplaying the fluctuating nature of strategic and political objectives, democratic debate, the ethics of war, and military proportionality, which statistical probabilistic AI reasoning is unable to replicate or simulate synthetically (Davis and Bracken 2022). For instance, by shifting the focus to technical prowess and precision in the conduct of war and away from discussions on whether war justifies the ends (i.e. from *jus ad bellum* to *jus in bello*), leaders risk eroding the critical link between the means and proportionality of war, and in turn, will tend to assume that tactical prescriptions are ethical and morally sound. In other words, AI's assumed tactical efficiency might provide leaders with an expedient *deus ex machina* to use military force divorced from consequences. For example, during the Second Iraq War, US Navy Captain Arthur Cebrowski stated that "network-enabled armies kill more of the right people quicker … *with fewer civilian casualties, warfare would be more ethical*" (Shachtman 2007, emphasis added). How might an AI system calculate what is a proportionate response? Who would be held responsible for the legal and ethical mistakes of machines?

Today, there is no reliable metric (legal, computational, or normative) to objectively measure disproportionate suffering (unethical, immoral, moral-injury related, superfluous, or excessive) during combat; it is ultimately subjective and requires human judgment. For example, "Bugsplat" software and its AI-ML enhanced successor SKYNET

have been developed and deployed by the US DoD to support human decision-makers in determining the most "appropriate" and "precise" payload for US drone strikes in order to destroy a target and calculate its impact (Sharkey 2014).[11] Scholars highlight the morally flawed design, exploitation, and the erosion of "ethical due care"[12] (Emery 2022, 182) in the use of technologies like these to conduct so-called "algorithmic assassinations" at a distance.[13] By contrast to traditional ("semi-autonomous") long-range strike systems (e.g. missile defense systems and unmanned drones), because AI-enabled autonomous weapons can select and engage their targets without humans intervention, soldiers are not only relieved of the moral gravity that accompanies the *experience* of killing but also the *decision* to kill (MacIntosh 2021).

Outsourcing parts or the entire ethical deliberation process to AI outputs – even if the ultimate decision to use lethal force remains with humans – will not answer the complex and subjective ethical-political dilemmas in war. Counterintuitively, ethical deliberation in chaoplexic war demands a degree of human subjective "inefficiency" – a core feature of ethical deliberation and democratic discourse – to face the challenges of complexity, contingency, and asymmetry in *jus in bello*, rather than encoded techno-rationalized "efficiency" (Derian 2000). Therefore, an undue focus on statistics (i.e. how many civilians have been killed or injured) and tactical efficacy risks eschewing critical debates on the initial rationale for war (*jus ad bellum*) or the altered character of war that technology like AI enables (Chamayou 2014).

Moreover, in uncertain and contingent contexts like war, subjective proportionality calculation is needed – to determine, for example, a genuine target in a warzone and weigh the target value by a probability of its presence and absence – for which AI statistical probabilistic inductive reasoning is arguably ill-equipped.[14] Therefore, integrating AI in the formulation of strategy in chaoplexic warfare – and the assessment of the role of military force within it – will require creativity, adaption, and an understanding of the likely consequences of any course of action, to exploit this uncertainty. As philosopher Pierre-Joseph Proudhon cautions: "the fecundity of the unexpected far exceeds the statesman's prudence" (Arendt 1970, 7).

Automation bias

What happens when military commanders place too much trust in AI systems? The shifting political economy and authoritative hierarchy of human-machine interactions can partly be attributed to the uncritical and often blind trust placed in intelligent machines, a cognitive affliction known as "automation bias."[15] This would describe situations where humans anthropomorphize machines and view technology as more capable than it is – ascribing human-like significance to AI – and thus use automation as a heuristic replacement for vigilant information seeking, cross-checking, and adequate processing supervision.[16] According to AI researcher Eliezer Yudkowsky, "anthropomorphic bias can be classed as insidious: it takes place with no deliberate intent, without conscious realization, and in the face of apparent knowledge" (Yudkowsky 2008). Consequently, people assume positive design intent even when presented with evidence of a system's failure. As the Scottish philosopher David Hume opined, "there is a universal tendency among mankind to conceive all beings like themselves … We find human faces in the moon, armies in the clouds" (Hume 1889, 11).

More worrisome, studies demonstrate that this phenomenon manifests itself in both experts and non-expert participants, and in military and civilian contexts. It cannot be mitigated by enhanced training protocols, and it can affect group and individuals decision-making processes equally (Skitka, Mosier, and Burdick 1998; Watson 2019). People's deference to machines can result a) from the presumption that machine decisions result from hard, empirically-based science; b) from the presumption that algorithms function at speeds and complexities beyond human capacity (the anthropomorphic argument), or c) from people's fear of being overruled or outsmarted by machines (Shekhtman 2016). Although few studies relate to this phenomenon in AI across multiple domains (e.g. driverless vehicles, medicine, aviation, and the financial markets), empirical studies demonstrate people's proclivity to automation bias (Wagner, Bornstein, and Howard 2018; Parasuraman and Manzey 2010). In the financial sector, for instance, the appeal of algorithmic "black box" trading is precisely due to the fact that it encourages automation bias – that is, the assumption that the trading recommendation of algorithms is far superior to those of humans (Pasquale 2016).

This phenomenon in a non-linear and contingent military context could mean that planners become more predisposed to view the judgments of AI as analogous (or even superior) to those of humans. An assumption that machines make better judgments than humans may also cause commanders to defer accountability and responsibility for the use of lethal force and thus neglect how AI in human-machine interactions is shaped by humans (e.g. algorithmic design, parameters, and settings that define the interactions, and how strategic objectives are defined and adapted), and thus how AI may influence decision-making in war (e.g. predictions and decision outputs, inductive reasoning, real-time situational awareness, and biases embedded in algorithms by their human creators). For example, a recent study of human–computer interaction in the US Air Force, with algorithms spanning over two decades, revealed a strong tendency to outsource judgment in using military force to machines and outsource accountability for the killing of non-combatants during warfare (Emery 2022). Outsourcing the practical judgment of human commanders to machines – in a flawed attempt to address the complex ethical-political dilemmas inherent in uncertain war – also risks undermining the moral assumptions that underpin the right to use lethal force in war (Renic 2019). Arendt notes that while there is a distinction between legal and normative moral issues, both presume the power of personal judgment and responsibility and are thus intrinsic to the "human condition" (Hill 2021, 163).

A recent study of surveillance in law enforcement highlights the effects of automation bias on predictive policing. The study notes: "The phenomenon of automation bias occurs in decision-making because humans tend to disregard or not search or contradictory information in light of a computer-generated solution that is accepted as correct" (Millar 2014, 122). This tendency is equally prevalent in fully automated decision-making systems and in cases where humans remain "on the loop" to make judgments on machine-decision outputs – or "mixed-mode" systems (Schwarz 2018a, 159). For example, the US Air Force's *Unmanned Aircraft Systems Flight Plan 2009–2047* outlines its vision for AI-enabled (data processing, software upgrades, sensor enhancements, etc.) autonomous drone swarm technology that will allow multiple drones to cooperate with a

variety of lethal and non-lethal ISR missions at the "command of a single pilot," i.e. "on the loop" (US Air Force 2009, 41).

Efforts to reduce the role of humans in the combat decisions prompted by tactical priorities will further diminish the role of commander's intent and permit life-and-death decisions to be made by algorithms who lack the intuition of humans – as well as other non-calculable human qualities such as compassion, respect, emotion, mercy – to sense something is wrong and change a course of action without explicit orders. Even where humans are "on the loop" monitoring the executions of certain decisions (e.g. targeting lists and launch and recall decisions), human decision-making would be too slow to react and thus intervene in AI-enabled warfare (Davis 2007).[17] A key takeaway of these studies is that despite human involvement in the decision-making process, automation bias can mean that errors and unethical or biased algorithmic decisions go undetected or unchallenged (Simonite 2019).[18]

Several analysts have warned that if human commanders place too much confidence in AI reasoning without fully understanding how machines reach a particular outcome, then decision-makers could trust machine-generated data implicitly and without scrutiny (Parasuraman and Riley 1997). For example, the Tesla Model 3 crash in 2018 – where a driver in autopilot mode plowed into a fire truck on a freeway – demonstrated the risk of placing too much trust in autonomous technology, though not necessarily in AI (Hawkins 2018). In this scenario, like other incidents involving vehicles in auto-pilot mode, the skill-based reasoning automated systems relying on "bottom-up" processing (or "skills-based reasoning") failed, and fatalities occurred because inattentive drivers did not realize that an automated driving assist feature still needs "top-down reasoning" – or human intuition, common sense, experience, and judgment (Cummings 2021).

The drive for predictability and centralized control over warfare and meaningful human control – that is, human-in-the-loop or human-on-the-loop, either positively to use lethal force or negatively to prevent an AI-guided accidental deployment – neglects the psychology of human-machine interactions, and above all, how automation bias shapes human-machine interaction in decision-making.[19] *In extremis*, this neglect might shift the hierarchy in the human-machine relationship from benign tools of our will and force-multipliers to becoming key influencers of ethical, political, and strategic concepts and norms through machine logic and rationality. In sum, the mystical quality of superhuman omniscience and omnipresence in AI overlords, acting on the uncertain and complex battlefield, will likely further diminish the role of humans in war.

Military techno-ethics and the moral responsibility of war

Can technology resolve the vexing ethical, moral, political dilemmas of war? The prevalent view that AI systems are morally objective and thus superior to (more "rational" than) human ("irrational") judgment and decision-making is becoming profoundly entrenched and presents us with a new range of psychological and ethical dilemmas; above all, deliberating on the costs and benefits of political-military objectives against unforeseeable but probable outcomes and experiencing the cognitive and emotional weight of those decisions (Bousquet 2018; Clark 2019). Coding ethics into AI-enabled capabilities has emerged as a possible solution to the complex, nuanced, and highly subjective ethical-political dilemmas of war: the political choices to use military force and the

psychological traumas of war combatants face in the conduct of war (Emery and Biggs 2022). Christopher Coker writes: "We are trying to 'moralize' weapons, an elegant term for abdicating control over our own ethical decision-making [to intelligent machines] … that may be better placed than use to make the right moral judgment" (Coker 2013, xxiii). Thus, the Panglossian quest to imbue AI with human conscience (or "AI consciousness") (Holland 2003) risks defusing moral responsibility of war to technology, "smoothing over" (rather than eliminating) moral and ethical tensions between discrimination, responsibility, and accountability for actions and accidents in war (Crawford 2013, 233).[20]

A bifurcated focus on the utopian or dystopian effects of AI and future warfare frequently neglects the intrinsic role of humans in the long casual chain associated with AI algorithms: the programmer, the designer, the military bureaucracy, military and political leaders, and the operator. Thus, it is challenging to mete out responsibility and blame for intentional unethical acts and war crimes or (human or machine) mishaps. This causal chain may also become socially constructed and anthropomorphized if the human creators define their self-worth and moral and ethical benchmarks in relation to the standards of superior, rational, and "efficient" machines. For example, proponents of lethal autonomous weapons such as drones frame them as efficient (less human cost) and effective (accurate discrimination of targets) tools of military force at a distance and as morally and ethically reliable (Foust 2013; Strawser 2010).

Roboticist Ronald Arkin argues that "intelligent robots can behave more ethnically in the battlefield than humans currently can" (Cornelia 2008). Similarly, AI researcher Gary Marcus posits that as driverless vehicles mature, they might become more moral than human drivers; *ipso facto*, the decision to drive a car would be inherently immoral (Marcus 2012). This perspective, however, neglects the potential of AI-enabled autonomous weapons to circumscribe the shared humanity that connects adversaries. Without the first-hand experience of the horrors of war, commanders may become overconfident, injudicious, and progressively desensitized to using AI-enabled autonomous weapons (Galliott 2016). Consequently, militaries that delegate decisions on the use of lethal force to machines might normalize (both combatant and potentially civilian) casualties and violence, thus potentially impeding any desire to prevent or terminate conflict (Vallor 2016). Ultimately, soldiers far removed from the battlefield might be incapable of developing fundamental military ethical virtues such as mercy, compassion, respect, and empathy (Vallor 2015) – a moral void machines cannot fill.[21] This is what is often called "moral de-skilling."

Elke Schwarz posits that AI coded ethics contains the "illusory promise of certainty, the fallacy of being able to offer a technical way of resolving ethical questions," thus obscuring the ethical questions (Schwarz 2018a, 198). For instance, the US DoD claims that "Bugsplat" can "produce a large body of *scientifically valid data, which enable weaponeers to predict the effectiveness* of weapons against most selected targets" (US Joint Chiefs 2013, emphasis added). Can humans retain their ethics and values if moral responsibility is outsourced to AI? And if so, might human ethics be programmed into AI and ultimately defined through algorithmic code, which we, due to automation bias, will use to justify lethal force?

Even if we accept the (albeit tenuous) argument that intelligent machines are morally and ethically preferable on the battlefield to humans, several problems remain. What

moral and ethical codes should we bake into AI's? Deontological, utilitarian, Kantian ethics, absolutism, virtue ethics, just war theory, divine command ethics, ethical due care, or something else? What might the effect of this choice be for human moral, legal, and personal responsibility and accountability in war? (Veruggio and Abney 2014).

Complicating this choice further, an ethical theory has not yet advanced to the point where broad agreement exists (even amongst moral philosophers) on the "correct" or "best" answer for the vast heterogeneity of moral circumstances, ethical conduct, and dilemmas (Anderson 2008). For instance, it cannot be assumed that there will ever be a satisfactory answer to the question of what should a person do in a type x ethical situation. Therefore, ethical military behavior, like many other domains, will be improvised based on experience, perceptual skills, and emotions rather than mastering an ethical-moral algorithmic code. In situations where disagreement existed over whether a particular action was morally acceptable, autonomous systems that are expected to behave ethically and morally would either be unable to act or make an arbitrary decision – neither of which would be satisfactory during combat.[22] Even if the goal of creating autonomous ethical machines were achieved, judgment would need to be made about the "moral status" of the non-human agent (i.e. vis-à-vis human agents).

Therefore, encoding ethical principles will require extensive analysis and deliberation before any precision can be achieved. As former DoD Chief of High-Value Targeting, Mark Garlasco, noted, "we cannot simply download international law [or ethics and morals] into a computer" (Singer 2009, 389). What do ethical principles mean in practice? For instance, does civilian support of the enemy in an insurgency make civilians a viable target? And how can torture be distinguished from mild pressure? Moreover, should designers encode these principles based on logical inductive statistical reasoning (i.e. proving and refuting theorems), or wait until the induction-deduction and top-down vs. bottom-up cognitive problems have been resolved, that is, when AI can produce reason by analogy and deductively reason like humans (Wallach and Allen 2009)?[23] As Austrian philosopher Ludwig Wittgenstein (and his successors) argued, concepts and principles such as morals and ethics contain structures deeply embedded in the social, cultural, and linguistic fabric of the human experience (Wittgenstein 1973). The burden of the challenge of answering these questions is, therefore, substantial.

Ronald Arkin argues that AI systems must appreciate the ethical significance of competing courses of action and use ethical and moral principles in a context-appropriate manner to address some of these vexing questions. Arkin's hypothetical solution is an algorithm that can obviate messy human emotion and irrational impulses and identify situations where there is a significant risk of unethical behavior and respond, either by restraining the system directly or alerting human operators who would intervene to resolve ethical dilemmas – a so-called "ethical governor" (Arkin 2009). For Arkin's ethical governor concept to be workable, however, ethics uploaded to an AI system must appreciate the nuances of competing courses of action and execute moral codes (not to kill unnecessarily, avoid collateral damage, not harming non-combatants, and adhering the Geneva and Hague Conventions, etc.) with high fidelity in a context-appropriate manner. "Ethical governors" would; therefore, either run the risk of machines acting unethically (e.g. contravening the principles, laws, and norms of war) and leave it to human operators to pick up the pieces, or require machines themselves to think, plan, and act "ethically," that is, through the optimization and verification of sensory

data with a judgment and choice of action equivalent to human capacity (Sparrow 2004).[24] Either way, the ability of humans to challenge the abstracted ethics of machines would be reduced, opening up a moral vacuum where neither laws, moral guides, or norms have adequate reach (Schwarz 2018b).

At a practical level, it is questionable whether AI-enabled capabilities can distinguish legitimate from illegitimate targets. To resolve this problem, some have argued that autonomous weapons should only be used in less complex strategic environments where there is a lower chance of encountering non-combatants. Critics stress the improbability of programming such context-specific and value-laden consequentialist reasoning principles (e.g. not to target civilian populations, what constitutes a legitimate combatant target, and the level of civilian casualties deemed acceptable) into algorithms (Hagerott 2014; Arkin 2009). At the very least, it remains an open empirical question whether machines can be trusted to implement the moral judgments of humans safely and reliably in warfare.

During asymmetric conflicts such as insurgencies, civil wars, and gray-zone conflicts such as the Russian invasion of Ukraine in 2022, reliable intelligence about whom to target needs to be based not only on situational awareness but also on attempts – albeit with limited hope of success – to determine the other side's intentions (i.e. the "theory of the mind") as a means to predict their future behavior in particular contingencies (Rauta and Stark 2022). The goal of coding human ethics into machines, such as Arkin's ethical governor – conceived on the assumption of the superiority of Cartesian rational ethical decisions lacking human emotion – understates the pivotal role emotion plays in ethical decision-making in war (Emery and Biggs 2022). As William James pointed out, human emotions are fundamentally enmeshed in every "rational" decision and ethical choice we make (James 2009). Therefore, human actions, emotions, and morals cannot be explained away by Spinozan general laws and then coded into algorithms (Tallis 2010). Valerie Morkevicius writes: "emotions can help us to act morally in four ways that are particularly relevant for the ethics of war … informing our moral intuition, generating empathy, and holding us accountable for our choices [that in turn] guide us towards more ethical behavior" (Morkevicius 2014, 10).

Although human judgment and prediction are far from perfect, evolutionary features of our social interactions (psychological, social, cultural, political, emotional, etc.) allow us to recognize subtle cues (e.g. facial expressions and emotions) that machines arguably cannot. These cues can prove critical, for example, in situations where lethal force is inappropriate, such as children being forced to carry empty guns, non-combatants attending to the wounded, or insurgents burying their dead (Sharkey 2014, 118). Because every ethical war decision is contextually and empirically bound, each choice and action have different ethical implications, thus making it virtually impossible to compute (Schwarz 2018a, 166–167). Consequently, if machine morality becomes anthropomorphized into existence by engineers using "fuzzy intuitions," the ethical coding enterprise will be reduced to merely a problem-solving exercise within constraints – or "operational morality" (Beavers 2010).

At a philosophical level, whether one adopts a consequentialist or a non-consequentialist view of the nature of the principles of *jus in bello* has important implications for whether and how AI meets the *jus in bello* requirement of discrimination. Philosopher Thomas Nagel argues that in warfare, people must acknowledge the "personhood" (or

humanity) of the enemy – *both* sides need to acknowledge that they are Kantian "ends in themselves" (Nagel 1972). Following Nagel's argument to its logical end – and without taking it purely at face value – until which time (arguably never) AI achieves the moral standing of "persons," they will be unable to meet the requirements of *jus in bello*.[25] Moreover, techno-ethics cannot account for cacophonic facets of the human condition such as empathy, emotion, intuition, compassion, revenge, remorse, cognitive bias, experience and learning, and many sensory perceptions that influence our decisions and shape our intentions and thus our responses to ethical and moral dilemmas (Simon 1987; Gross 2002; Gayer et al. 2009; Mehta, Jones, and Josephs 2008; Gladue, Boechler, and McCaul 1989). Human judgment, ethics, and beliefs are conditioned by our cognitive capacities and limitations that give us a sense of reason and deliberation (i.e. human agency), which has evolved to cope with the breakdown of rational control (Veruggio and Abney 2014, 355–356).[26] Schwarz writes that we should not attempt "to normalize and homogenize something [human ethics and morals] that can be neither normalized or homogenized owing to its inherent contingent nature" (Schwarz 2018a, 186).

Conclusion

In her political theory magnum opus, *The Human Condition*, Hannah Arendt sought to consider the human condition "from the vantage point of *our newest experience* and *our most recent fears* … in order to *think though what we are doing*" (Arendt 1998, 5, emphasis added). In the spirit of inquiry inspired by Arendt's work, this article has addressed the vexing and complex issues that arise as humans become increasingly intertwined with intelligent machines. Above all, the fear that by absolving human decision-makers of the political and ethical dilemmas and paradoxes of war by outsourcing these tasks to machines, we risk creating moral vacuums that hollow out meaningful ethical and moral deliberation in the illusory quest of riskless, "rational," and frictionless war. Moreover, the fallacy that technology such as AI – as the latest manifestation of the scientific way of war – offers a panacea to the uncertainty and contingency of war removes us one causal (physical and psychological) step further from war-making, making us less equipped to control, or even foresee, the consequences of war as a fundamentally "human thing". The aura of tactical efficiency associated with AI provides leaders with an expedient *deus ex machina* to use military force divorced from consequences.

This article has provided a counterpoint to the argument prevalent today in the defense community that delegating war-making to AI offers a viable solution to the psychological and biological fallibility of humans in war while simultaneously retaining meaningful human control over the AI-powered war machine. It argues that this latter argument neglects the psychological processes of human-machine integration – or human technology symbiosis – specifically, how AI human-machine interactions are both shaped by humans and shape military decision-making. Because of the complex and quintessentially cognitive-psychological symbiosis of man and machine – across the tactical to the strategic decision-making continuum – algorithms cannot be merely passive neutral force multipliers of advanced capabilities. Instead, as deep human-machine symbiosis will alter and shape the psychological mechanisms that make us

who we are, AI agents, as they learn and evolve, will likely become – either inadvertently or more probable by conscious choice – *de facto* strategic actors in war.

The article contributes three key psychological insights that consider human-machine interactions and political-ethical dilemmas in future AI-enabled warfare. These insights elucidate the *de facto* AI commander problem advanced in the paper.

First, efforts to make war faster, more lethal, asymmetric, and efficient are being accomplished through a socio-technical psychological process of human-machine integration. Human-machine integration is part of a broader evolutionary dovetailing of humanity and technology. AI represents a new manifestation of the pursuit of a technological solution, a new manifestation of an omniscient technological solution to the ethical-political dilemmas of war. The logical end of this trajectory is an AI commander. The danger is that decision-makers may seek to reconcile the paradox of war by outsourcing our consciences in the use of lethal force to non-human agents who are ill-equipped to fill this ethical, moral, and void.

Second, AI is the newest means commanders leverage in their technical-scientific quest to impose predictability and certainty in chaotic and complex contemporary warfare (or chaoplexic warfare). Until AI can produce testable hypotheses or reason by analogy and deductively reason like humans, they will, however, be incapable of understanding the real world, and the role of human agents in "mission command" will be even more critical in future AI-enabled warfare. Moreover, specific cognitive biases (illusion of control, heuristic shortcuts, automation bias) associated with human-machine interactions can compound the "illusion of control" problem.

The research also found that biases (1) can make decision-makers prone to use capabilities just because they invested time and resources in their acquisition, which may produce false positives about the necessity for war (the Einstellung effect); (2) can make decision-makers prone to unreflectively assign positive moral attributes to the latest techno-military Zeitgeist; and (3) can lead to humans anthropomorphizing machines and viewing technology as a heuristic replacement for vigilant information seeking, cross-checking, and adequate processing supervision (automation bias).

Finally, the notion that human ethics can be coded into AI algorithms as a possible solution to war's subjective and multifaceted ethical-political dilemmas is technically, theoretically, ontologically, and psychologically problematic, and ethically and morally questionable. These vexing questions require open and broad democratic debate and multi-disciplinary deliberation (NATO OTAN 2020; Schmitt 2013; Ekelhoff and Paoli 2019; Tarraf 2019; ICRC 2018). Abdicating control over our ethical decision-making to machines – under the assumption that AI can make superior moral judgments – risks defusing (rather than eliminating) the moral responsibility of war to technology.

AI opens up a Pandora's box of images, illusions, myths, hopes, and fears that humanity since the ancient Greeks have impressed upon artificial life, automata, self-moving devices, and human enhancements (Mayor 2020). Because the psychological issues associated with the human-machine symbiosis borrow from many other related disciplines (military and applied ethics, law, philosophy, neuroscience, computer science, etc.), the challenge of unsnarling old from new issues in the drive for analytical erudition and moral clarity is substantial. The current transition to a new era of autonomous "intelligent" machines has its providence in ancient times and

thus offers us a valuable opportunity to ruminate about not only about what these advances will mean for the character of future war, but also what it means for the human condition. The "AI commander" notion expounded here not only serves as both a warning about the potential consequences of neglecting the salience of human psychology in human-machine interactions, but also provides us with the kinds of deep questions we need to consider as this critical symbiosis between man and machine evolves in its latest incarnation in the age of AI.

Notes

1. There are three different types of AI: artificial "narrow" intelligence (ANI), artificial general intelligence (AGI) that matches human levels of intelligence, and artificial superintelligence (ASI) (or "superintelligence") that exceeds human intelligence. The debate about when and whether AGI (let alone ASI) will emerge is highly contested and thus inconclusive. This article focuses on task-specific "narrow AI" (or "weak AI"), which is rapidly diffusing and maturing (e.g., facial recognition, natural language processing, navigation, and digital assistants such as Siri and Google Assistant).
2. Many AI-enabled weapons have already been deployed, are already operational, or have already been developed by militaries. Examples include: Israel's autonomous Harpy loitering munition; China's "intelligentized" cruise missiles, AI-enhanced cyber capabilities, and AI-augmented hypersonic weapons; Russia's armed and unarmed autonomous unmanned vehicles and robotics; and US "loyal wingman" human-machine teaming (unmanned F-16 with a manned F-35 or F-22) program, intelligence, surveillance, and reconnaissance (ISR) space-based systems, and various AI-ML-infused command and control support systems.
3. The notion of "meaningful human control" in the context of autonomous weapons, while gaining currency is also contested; see Moyes 2016.
4. The pursuit of riskless war – dictated by the exigencies of war – is logical and morally laudable. This same drive, however, is moving us ever closer to a mode of violence that strains the basic foundations upon which our moral justifications for killing in war rest.
5. Mechanistic dehumanization creates an "emotional disengagement" by engaging brain parts that give rise to an "emotionally dysfunctional cognitive mode."
6. Whether values and morals are independent of rationality (advocated by David Hume) or correlated with rationality (advocated by Immanuel Kant) remains an unresolved philosophical question. The Humean view is generally more common in discussions about AI that assume machine "values" (i.e., goals and motives) are independent of their "rationality" (i.e., their reasoning about how to accomplish goals), see Chalmers 2010.
7. For example, throughout the nineteenth century aerial bombing placed a significant strain on the utilitarian logic of military necessity; see Smith 2022.
8. For example, the 1994 friendly fire shootdown of two US Army Blackhawks in the no-fly zone over Northern Iraq was caused by human misidentification of *the Airborne Warning and Control System* (AWAC), and the 1988 USS Vincennes mistaken attack of an Iranian civilian airline was due to human error caused by errors in the Aegis missile defense system human-machine interface.
9. AI-ML techniques (e.g., image recognition, pattern recognition, and natural language processing) inductively (inference from general rules) fill the gaps in missing information to identify patterns and trends, thereby increasing the speed and accuracy of certain standardized military operations, including open-source intelligence collation, satellite navigation, and logistics.
10. Algorithms are generally trained once on a data-set, thus making them incapable of learning new information without retraining. By contrast, the human brain learns constantly using knowledge gained over time and building on it as it encounters new information and environments.

11. The role of AI in an *advisory* or *decision-support* capacity – while the underlying technology may be similar – is very different from AI in decision-making or active battlefield capacity. While the former might, in theory, offer commanders and their tactical leaders improved battlefield awareness and thus resulting in more informed decisions, because of the various cognitive psychological issues described, AI may perform an outsize role, thereby obviating any potential ethical advantages of this synthesis. For a more detailed decision of this slippery slope argument, see Johnson 2022.
12. "Ethical due care" refers to a positive commitment to save civilian lives, not merely to apply the rule of proportionality and kill no more civilians than is militarily necessary (see Emery 2022).
13. Other examples of "algorithmic assassinations" include Israeli's Harpy "fire and forget" autonomous drone, which uses its software to prioritize the threat and then verify the target; and the US Phalanx system that "automatically detects, evaluates, tracks, and engages," and performs kill assessments against its targets. Both these "autonomous" systems rely on pre-programmed data-sets and are incapable of self-learning or targeting objects that do not emit threat signals or trajectories that can be sensed.
14. A proportionality calculation should be based on an estimate of the potential differences in military outcome if a particular action has not been taken.
15. Evidence suggests that peoples' levels of trust in machines can also be affected by other factors, including culture, familiarity, and anthropomorphizing (Butcher 2022).
16. The source of excessive trust and automation bias might also be a machine's (real or perceived) capability, regardless of anthropomorphism.
17. Whether future AI-enabled systems will be able to match or surpass the performance of humans "in-the loop" is an open empirical question.
18. For example, there is mounting evidence that AI data-sets can perpetuate biased decision-making in the medical domain, such as prioritizing health care for white patients over black ones.
19. The control "loop" concept has also been criticized for the crude distinction it makes about decision-making. Command decision-making can be both immediate (i.e., in an individual targeting situation) or more broadly defined (i.e., associated with algorithmic programming); see Heyns 2016.
20. The concept of "AI consciousness" (or "machine consciousness" is a heavily contested and vexing neurological, philosophical, and psychological issue.
21. As a counterpoint, recent evidence suggests that drone operators do not feel the emotionally distance – and arguably its subsequent dehumanization – but quite the opposite; they are allowed, due to modern technology, to come psychologically much closer to the enemy, resulting in cases of moral-injury, PTSD, and other mental afflictions for drone operators as much as for combatants on the physical battlefield (Chappelle et al. 2014).
22. Some non-ethicist philosophers believe that in cases where no agreement exists on ethical dilemmas a metaethical concept known as ethical relativism can be applied – that is, the view that both sides are correct and there are no absolutes wrongs. This approach is generally rejected by ethicists.
23. For a discussion about the need to merge the cognitive top-down with the bottom-up approach, see Wallach and Allen 2009.
24. Bioethicist Robert Sparrow adopted Alan Turing's 1950 "Turing Test" to serve as a basis for testing whether machines had achieved the moral standing of humans: the hypothetical "The Turing Triage Test."
25. The ability of machines to think for themselves, and broader questions about the moral standing of machines, would require the creation of AGI (or "strong AI") – the ability of machines to understand and learn any task that a human can. AGI would bring about AI systems able to reason, plan, learn, represent knowledge, and communicate in natural language.
26. In most ethical traditions, "agency" matters because of our "theory of the mind." Human social evolution has made us hardwired to attribute agency – and thus intention and blame – promiscuously to those we interact with.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

James Johnson is a Lecturer in Strategic Studies at the University of Aberdeen. He is also an Honorary Fellow at the University of Leicester, a Non-Resident Associate on the ERC-funded Towards a Third Nuclear Age Project, and a Mid-Career Cadre with the Center for Strategic Studies (CSIS) Project on Nuclear Issues. He is the author of *Artificial Intelligence and the Future of Warfare: USA, China & Strategic Stability* (MUP, 2021), and *AI & the Bomb: Nuclear Strategy and Risk in the Digital Age* (OUP, 2023).

References

Acton, James M. 2020. "Cyber Warfare & Inadvertent Escalation." *Daedalus* 149 (2): 133–149. doi:10.1162/daed_a_01794

Aircraft Accident Investigation Board Report. 1994. *US Army UH-60 Blackhawk Helicopters 87–26000 and 88–26060*, vol. 1 (Executive Summary), May 27. Washington, DC: US Department of Defense.

Alberts, David S., and Richard E. Hayes. 2003. *Power to the Edge: Command and Control in the Information Age*. Washington, DC: Department of Defense Command and Control Research Program.

AlphaStar Team. 2019. "Alphastar: Mastering the Real-Time Strategy Game Starcraft II." *DeepMind Blog*, January 24. Accessed December 22, 2022. https://www.deepmind.com/blog/alphastar-mastering-the-real-time-strategy-game-starcraft-ii.

Anderson, Susan. 2008. "Asimov's 'Three Laws of Robotics,' and Machine Metaethics." *AI and Society* 22 (4): 477–493. doi:10.1007/s00146-007-0094-5

Arendt, Hannah. 1970. *On Violence*. New York: Harcourt.

Arendt, Hannah. 1998. *The Human Condition*. Chicago: University of Chicago Press.

Arkin, Ronald. 2009. *Governing Lethal Behavior in Autonomous Robots*. Boca Raton: Chapman.

Arquilla, John, and David Ronfeldt. 1997. "Looking Ahead: Preparing for Information-age Conflict." In *In Athena's Camp: Preparing for Conflict in the Information Age*, edited by John Arquilla, and David Ronfeldt, 439–501. Santa Monica: RAND.

Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. 2014. "Neural Machine Translation by Jointly Learning to Align and Translate." *ArXiv*, article # 1409.0473.

Beavers, Anthony. 2010. "Editorial." *Ethics & Information Technology* 12 (3): 207–208. doi:10.1007/s10676-010-9244-4

Bengio, Yoshua, Yann Lecun, and Geoffrey Hinton. 2021. "Deep Learning for AI." *Communications of the ACM* 64 (7): 58–65. doi:10.1145/3448250

Beyerchen, Alan. 1992–1993. "Clausewitz, Nonlinearity, and the Unpredictability of War." *International Security* 17 (3): 59–90. doi:10.2307/2539130

Boulanin, Vincent, ed. 2020. *Artificial Intelligence, Strategic Stability and Nuclear Risk*. Stockholm: SIPRI.

Bousquet, Antoine. 2008. "Chaoplexic Warfare or the Future of Military Organization." *International Affairs* 84 (5): 915–929. doi:10.1111/j.1468-2346.2008.00746.x

Bousquet, Antoine. 2018. *The Eye of War: Military Perception from the Telescope to the Drone*. Minneapolis: University of Minnesota Press.

Brough, Michael W. 2007. "Dehumanization of the Enemy and the Moral Equality of Soldiers." In *Rethinking the Just War Tradition*, edited by Michael W. Brough, John W. Lango, and Harry van der Linden, 149–167. New York: SUNY Press.

Butcher, Fiona D. 2022. "Psycho-social Factors Influencing Trust in Artificial Intelligence Advice Systems." Thesis. Leicester: University of Leicester. doi:10.25392/leicester.data.20310096.v1.

Caron, Jean-François. 2020. *Contemporary Technologies and the Morality of Warfare: The War of the Machines*. London: Routledge.
Castel, Robert. 1991. "From Dangerousness to Risk." In *The Foucault Effect; Studies in Governmentality*, edited by Graham Burchell, Colin Gordon, and Peter Miller, 281–298. Hertfordshire: Harvester Wheatsheaf.
Cebrowski, Arthur K. 1999. "Sea, Space, Cyberspace: Borderless Domains." Lecture Delivered to the US Naval College, Newport, RI, February 26.
Chalmers, David. 2010. "The Singularity: A Philosophical Analysis." *Journal of Consciousness Studies* 17: 7–65. doi:10.1002/9781118922590.ch16.
Chamayou, Gregorie. 2014. *A Theory of the Drone, Translated by Janet Lloyd*. New York: New Press.
Chappelle, Wayne L., Kent D. McDonald, Lillian Prince, Tanya Goodman, Bobbie N. Ray-Sannerud, and William Thompson. 2014. "Symptoms of Psychological Distress and Post-Traumatic Stress Disorder in United States Air Force 'Drone' Operators." *Military Medicine* 179: 63–70. doi:10.7205/MILMED-D-13-00501
Clark, Andy. 2003. *Natural Born Cyborgs, Technology & Future of Human Intelligence*. Oxford: Oxford University Press.
Clark, Lindsay. 2019. *Gender and Drone Warfare: A Hauntological Perspective*. London: Routledge.
Cohen, Eliot. 1996. "A Revolution in Warfare." *Foreign Affairs* 75 (2): 34–54.
Coker, Christopher. 2008. *Ethics and War in the 21st Century*. London: Routledge.
Coker, Christopher. 2013. *Warrior Geeks: How 21st-Century Technology Is Changing the Way We Fight and Think About War*. London: Hurst & Company.
Cornelia, Dean. 2008. "A Soldier, Taking Orders from Its Ethical Judgment Center." *The New York Times*, November 24. Accessed December 21, 2022. https://www.nytimes.com/2008/11/25/health/25iht-25robots.18126102.html.
Crawford, Neta C. 2013. "Bugsplat: Us standing Rules of Engagement, International Humanitarian law, Military Necessity, and non-Combatant Immunity." In *Just War: Authority, Tradition, and Practice*, edited by Anthony Lang, 397–422. Washington, DC: Georgetown University Press.
Creveld, Martin Van. 2003. *Command in War*. Cambridge, MA: Harvard University Press.
Cummings, Mary L. 2021. "Rethinking the Maturity of Artificial Intelligence in Safety-Critical Settings." *AI Magazine* 42 (1): 6–15.
Davis, Daniel. 2007. "Who Decides: Man or Machine?" *Armed Forces Journal*, November 1. Accessed December 20, 2022. http://armedforcesjournal.com/who-decides-man-or-machine/.
Davis, Paul K., and Paul Bracken. 2022. "Artificial Intelligence for Wargaming and Modeling." *The Journal of Defense Modeling and Simulation*. Online First, February 8.
Derian, James Der. 2000. "Virtuous War/Virtual Theory." *International Affairs* 766 (4): 772–788. doi:10.1111/1468-2346.00164.
Dobos, Ned. 2020. *Ethics, Security, and The War-Machine: The True Cost of the Military*. Oxford: Oxford University Press.
Eidelman, Scott, Christian S. Crandall, and Jennifer Pattershall. 2009. "The Existence Bias." *Journal of Personality and Social Psychology* 97 (5): 765–775. doi:10.1037/a0017058
Ekelhoff, Merel, and Giacomo Persi Paoli. 2019. "The Human Element in Decisions About The Use of Force." Geneva: UNIDIR. Accessed December 15, 2022. https://unidir.org/sites/default/files/2020-03/UNIDIR_Iceberg_SinglePages_web.pdf.
Emery, John R. 2022. "Probabilities Towards Death: Bugsplat, Algorithmic Assassinations, and Ethical Due Care." *Critical Military Studies* 8 (2): 179–197. doi:10.1080/23337486.2020.1809251
Emery, John R, and Hadley Biggs. 2022. "Human, All Too Human: Drones, Ethics, and the Psychology of Military Technologies." *Political Psychology* 43 (3): 605–613. doi:10.1111/pops.12809
Foust, Joshua. 2013. "The Liberal Case for Drones." *Foreign Policy*, May 13.
French, Shannon E., and Anthony I. Jack. 2015. "Dehumanizing the Enemy: The Intersection of Neuroethics and Military Ethics." In *Responsibilities to Protect: Perspectives in Theory and Practice*, edited by David Whetham, and Bradley J. Strawser, 165–195. Leiden: Brill.

Galliott, Jai. 2016. "War 2.0: Drones, Distance and Death." *International Journal of Technoethics* 7 (2): 61–76. doi:10.4018/IJT.2016070104

Gayer, Corinna Carmen, Shiri Landman, Eran Halperin, and Daniel Bar-Tal. 2009. "Overcoming Psychological Barriers to Peaceful Conflict Resolution: The Role of Arguments About Losses." *Journal of Conflict Resolution* 53 (6): 951–975. doi:10.1177/0022002709346257

Gill, Amandeep Singh. 2019. "Artificial Intelligence and International Security: The Long View." *Ethics and International Affairs* 33 (2): 169–179. doi:10.1017/S0892679419000145

Gladue, Brian A., Michael Boechler, and Kevin D. McCaul. 1989. "Hormonal Response to Competition in Human Males." *Aggressive Behavior* 15 (6): 409–422. doi:10.1002/1098-2337(1989)15:6<409::AID-AB2480150602>3.0.CO;2-P

Goldfarb, Avi, and Jon Lindsay. 2022. "Prediction and Judgment: Why Artificial Intelligence Increases the Importance of Humans in War." *International Security* 46 (3): 7–50. doi:10.1162/isec_a_00425

Gross, Janice J. 2002. "Emotion Regulation: Affective, Cognitive, and Social Consequences." *Psychophysiology* 39 (3): 281–291. doi:10.1017/S0048577201393198

Hagerott, Mark. 2014. "Lethal Autonomous Weapons Systems from a Military Officer's Perspective: This Time is Different: Offering a Framework and Suggestions." Paper presented at the United Nations Informal Meeting of Experts at the Convention on Conventional Weapons, May 15, Geneva, Switzerland.

Haslam, Nick. 2006. Dehumanization: An Integrative Review." *Personality and Social Psychology Review* 10 (3): 252–264. doi:10.1207/s15327957pspr1003_4

Hawkins, Andrew J. 2018. "Tesla Model S Plows into a Fire Truck while Using Autopilot." *The Verge*, January 23.

Heyns, Christof. 2013. "Report of the Special Rapporteur on Extrajudicial, Summary or Arbitrary Executions, United Nations Human Rights Council." A/HRC/23/47.

Heyns, Christof. 2016. "Autonomous Weapons Systems: Living a Dignified Life and Dying a Dignified Death." In *Autonomous Weapons Systems: Law, Ethics, Policy*, edited by Nehal Bhuta, Susanne Beck, and Robin Geiss, 3–20. Cambridge: Cambridge University Press.

Hill, Samantha R. 2021. *Critical Lives: Hannah Arendt*. New York: Reaktion Books.

Holland, Owen. 2003. *Machine Consciousness*. New York: Imprint Academic.

Horowitz, Michael C. 2010. *The Diffusion of Military Power: Causes and Consequences for International Politics*. Princeton: Princeton University Press.

Howard, Jennifer. 2011. "Jefferson Lecture, Drew Faust Traces the Fascination of War, From Homer to Bin Laden." *The Chronicle of Higher Education*, May 2. Accessed December 20, 2022. https://www.chronicle.com/article/in-jefferson-lecture-drew-faust-traces-the-fascination-of-war-from-homer-to-bin-laden/.

Human Rights Watch. 2012. "Losing Humanity: The Case against Killer Robots." November. Accessed December 15, 2022. https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots.

Hume, David. 1889. *The Natural History of Religion*. London: A. and H. Bradlaugh Bonner.

Hume, David. 1992. *A Treatise of Human Nature*. New York: Prometheus.

ICRC. 2018. *Ethics and Autonomous Weapon Systems: An Ethical Basis for Human Control?* Geneva: International Committee of the Red Cross.

James, William. 2009. *Varieties of Religious Experience: A Study of Human Nature*. New York: Signet.

Jervis, Robert. 1978. "Cooperation Under the Security Dilemma." *World Politics* 30 (2): 167–214. doi:10.2307/2009958

Johnson, James. 2021. "The End of Military-Techno Pax Americana? Washington's Strategic Responses to Chinese AI-Enabled Military Technology." *The Pacific Review* 34 (3): 351–378. doi:10.1080/09512748.2019.1676299

Johnson, James. 2022. "Automating the OODA Loop in the Age of Intelligent Machines: Reaffirming the Role of Humans in Command-and-Control Decision-Making in the Digital Age." *Defence Studies*, online first. doi:10.1080/14702436.2022.2102486.

Kahn, Herman. 1965. *On Escalation: Metaphors and Scenarios*. Cambridge, MA: Harvard University Press.
Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Penguin.
Kahneman, Daniel, and Jonathan Renshon. 2009. "Hawkish Biases." In *American Foreign Policy and the Politics of Fear: Threat Inflation Since 9/11*, edited by Trevor Thrall, and Jane Kramer, 79–96. New York: Routledge.
Keegan, John. 1976. *The Face of Battle*. London: Cape.
Kissinger, Henry, Eric Schmidt, and Daniel Huttenlocher. 2021. *The Age of AI and Our Human Future*. London: John Murray.
Kleinig, John. 2015. 'What's All the Fuss with Police Militarization?" *The Critique*, March 17.
Kramer, Eric-Hans. 2015. "Mission Command in the Information Age: A Normal Accidents Perspective on Networked Military Operations." *Journal of Strategic Studies* 38 (4): 445–466. doi:10.1080/01402390.2013.844127
Lerner, Jennifer S., and Dacher Keltner. 2001. "Fear, Anger, and Risk." *Journal of Personality and Social Psychology* 81 (1): 146–159. doi:10.1037/0022-3514.81.1.146
Lieber, Keir. 2000. "Grasping the Technological Peace: The Offense-Defense Balance and International Security." *International Security* 25 (1): 71–104. doi:10.1162/016228800560390
MacIntosh, Duncan. 2015. "PTSD Weaponized: A Theory of Moral Injury." Paper presented at the Center for Ethics and the Rule of Law at the University of Pennsylvania Law School, December 3–5.
MacIntosh, Duncan. 2021. "Fire and Forget: A Moral Defense of the Use of Autonomous Weapons Systems in War and Peace." In *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare*, edited by Jai Galliott, Jens Ohlin, and Duncan MacIntosh, 9–23. Oxford: Oxford University Press.
Marcus, Gary. 2012. "Moral Machines." *The New Yorker*, November 24.
Mayor, Adrienne. 2020. *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*. Princeton: Princeton University Press.
Mehta, Pranjal H., A. C. Jones, and R. A. Josephs. 2008. "The Social Endocrinology of Dominance: Basal Testosterone Predicts Cortisol Changes and Behavior Following Victory and Defeat." *Journal of Personality and Social Psychology* 94 (6): 1078–1093. doi:10.1037/0022-3514.94.6.1078
Meijer, Albert, and Martijn Wessels. 2019. "Predictive Policing: Review of Benefits and Drawbacks." *International Journal of Public Administration* 42 (12): 1031–1039. doi:10.1080/01900692.2019.1575664
Millar, Kevin. 2014. "Total Surveillance, Big Data, and Predictive Crime Technology: Privacy's Perfect Storm." *Journal of Technology Law & Policy* 19 (1): 106–145.
Morkevicius, Valarie. 2014. "Tin Men: Ethics, Cybernetics and the Importance of Soul." *Journal of Military Ethics* 13 (1): 3–19. doi:10.1080/15027570.2014.908011
Moyes, Richard. 2016. "Meaningful Human Control." In *Lethal Autonomous Weapons Systems: Technology, Definition, Ethics, Law & Security*, edited by Robin Geiss, and Henning Lahmann, 239–249. Berlin: Federal Foreign Office.
Nagel, Thomas. 1972. "War and Massacre." *Philosophy & Public Affairs* 1 (2): 123–144. doi:10.1177/000276427201500678.
NATO OTAN. 2020. *Science & Technology Trends 2020–2040: Exploring the S&T Edge*. Brussels: NATO Science & Technology Organization.
Norvig, Peter. 2014. *Artificial Intelligence: A Modern Approach*. 3rd ed. Harlow: Pearson Education.
O'Creevy, Mark Fenton, Nigel Nicholson, Emma Soane, and Paul Willman. 2003. "Trading on Illusions: Unrealistic Perceptions of Control and Trading Performance." *Journal of Occupational and Organizational Psychology* 76 (1): 53–68. doi:10.1348/096317903321208880
O'Hanlon, Michael E. 2018. "A Retrospective on the So-called Revolution in Military Affairs, 2000-2020." *Brookings*, September. Accessed December 14, 2022. https://www.brookings.edu/research/a-retrospective-on-the-so-called-revolution-in-military-affairs-2000-2020/8.

Parasuraman, Raja, and Dietrich Manzey. 2010. "Complacency and Bias in Human Use of Automation: An Attentional Integration." *Human Factors* 52 (3): 381–410. doi:10.1177/0018720810376055

Parasuraman, Raja, and Victor Riley. 1997. "Humans and Automation: Use, Misuse, Disuse, Abuse." *Human Factors* 39 (2): 230–253. doi:10.1518/001872097778543886

Paret, Peter. 1985. *Clausewitz, and the State: The Man, His Theories and His Times*. Princeton: Princeton University Press.

Pasquale, Frank. 2016. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Pawlyk, Oriana. 2020. "Rise of the Machines: AI Algorithm Beats F-16 Pilot in Dogfight." *Military.com*, August 24. Accessed December 22, 2022. https://www.military.com/daily-news/2020/08/24/f-16-pilot-just-lost-algorithm-dogfight.html.

Rauta, Vladimir, and Alexandra Stark. 2022. "What Does Arming an Insurgency in Ukraine Mean?" *Lawfare*, April 3.

Renic, Neil. 2019. "Justified Killing in an Age of Radically Asymmetric Warfare." *European Journal of International Relations* 25 (2): 408–430. doi:10.1177/1354066118786776

Roff, Heather M. 2014. "The Strategic Robot Problem: Lethal Autonomous Weapons in War." *Journal of Military Ethics* 13 (3): 211–227. doi:10.1080/15027570.2014.975010

Russell, Stuart. 2019. *Human Compatible*. New York: Viking Press.

Sauer, Frank. 2021. "How (Not) to Stop the Killer Robots: A Comparative Analysis of Humanitarian Disarmament Campaign Strategies." *Contemporary Security Policy* 42 (1): 4–29. doi:10.1080/13523260.2020.1771508

Schmitt, Michael. 2013. "Autonomous Weapons Systems and International Humanitarian Law: A Reply to the Critics." *Harvard National Security Journal*, online edition. Accessed December 22, 2022. https://bit.ly/3ip5pyh.

Schwarz, Elke. 2018a. *Death Machines: The Ethics of Violent Technologies*. Manchester: Manchester University Press.

Schwarz, Elke. 2018b. "Technology and Moral Vacuums in Just War Theorising." *Journal of International Political Theory* 14 (3): 280–298. doi:10.1177/1755088217750689

Shachtman, Noah. 2007. "How Technology Almost Lost the War: In Iraq, the Critical Networks are Social – Not Electronic." *Wired*, November 27.

Sharkey, Noel. 2014. "Killing Made Easy: From Joystick to Politics." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Patrick Lin, Keith Abney, and George Bekey, 111–129. Cambridge, MA: MIT Press.

Shekhtman, Lonnie. 2016. "Why Do People Trust Robot Rescuers More Than Humans?" *Christian Science Monitor*, March 1.

Simon, Herbert A. 1987. "Making Management Decisions: The Role of Intuition and Emotions." *The Academy of Management Executive* 1 (1): 57–64. doi:10.4018/978-1-5225-0731-4.ch018.

Simonite, Tom. 2019. "A Health Care Algorithm Offered Less Care to Black Patients." *Wired*, October 24.

Singer, Peter. 2009. *Wired for War: The Robotics Revolution and Conflict in the 21st Century*. New York: Penguin.

Skitka, Linda J., Kathleen Mosier, and Mark Burdick. 1998. "Automation Bias: Decision Making and Performance in High-Tech Cockpits." *International Journal of Aviation Psychology* 8 (1): 47–63. doi:10.1207/s15327108ijap0801_3

Smith, Brian. 2022. *A History of Military Morals: Killing the Innocent*. Leiden: Brill.

Sparrow, Robert. 2004. "The Turing Triage Test." *Ethics & Information Technology* 6 (4): 203–213. doi:10.1007/s10676-004-6491-2

Strawser, Bradley J. 2010. "Moral Predators: A Duty to Employ Unmanned Aerial Vehicles." *Journal of Military Ethics* 9 (4): 342–368. doi:10.1080/15027570.2010.536403

Tallis, Raymond. 2010. *Aping Mankind: Neuromania, Darwinists, and the Misrepresentation of Humanity*. New York: Atlantic Books.

Talmadge, Caitlin. 2019. "Emerging Technology and Intra-War Escalation Risks: Evidence from the Cold War, Implications for Today." *Journal of Strategic Studies* 42 (6): 864–887. doi:10.1080/01402390.2019.1631811

Tarraf, Danielle, et al. 2019. *The Department of Defense Posture for Artificial Intelligence: Assessment and Recommendations*. Santa Monica: RAND Corporation.

Thompson, Suzanne C. 1999. "Illusions of Control: How We Overestimate Our Personal Influence." *Current Directions in Psychological Science* 8 (6): 187–190. doi:10.1111/1467-8721.00044

Thoreau, Henry David. 1906. *Civil Disobedience in The Writings of Henry David Thoreau, Vol. 4*. Boston: Houghton Mifflin.

Thucydides. 1996. *The Landmark Thucydides: A Comprehensive Guide to the Peloponnesian War*, edited by Robert B. Strassler, translated by Richard Crawley. New York: Free Press.

US Air Force. 2009. *Unmanned Aircraft Systems Flight Plan 2009-2047*. Washington, DC: US Air Force Headquarters.

US Department of Defense. 2022. *Summary of the Joint All-Domain Command and Control (JADC2) Strategy*. Washington, DC: US Department of Defense.

US Joint Chiefs. 2013. *Joint Publication 3–60: Joint Targeting*. Accessed December 15, 2022. https://www.justsecurity.org/wp-content/uploads/2015/06/Joint_Chiefs-Joint_Targeting_20130131.pdf.

US Office of Naval Research. 2014. *Data Focused Naval Tactical Cloud (DF-NTC)*. ONR Information Package. Accessed December 15, 2022. https://docplayer.net/23690522-Data-focused-naval-tactical-cloud-df-ntc-onr-information-package.html.

Vallor, Shannon. 2015. "Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character." *Philosophy & Technology* 28 (1): 107–124. doi:10.1007/s13347-014-0156-9

Vallor, Shannon. 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. Oxford: Oxford University Press.

Veruggio, Gianmarco, and Keith Abney. 2014. "The Applied Ethics for a New Science." In *Robot Ethics: The Ethical and Social Implications of Robotics*, edited by Lin Patrick, Keith Abney, and George Bekey, 347–364. Cambridge, MA: MIT Press.

Wagner, Alan R., Jason Bornstein, and Ayanna Howard. 2018. "Computing Ethics: Overtrust in the Robotics Age." *Communications of the ACM* 61 (9): 22–24. doi:10.1145/3241365

Wallach, Wendell, and Colin Allen. 2009. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.

Watson, David. 2019. "The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence." *Minds and Machines* 29: 417–440. doi:10.1007/s11023-019-09506-6

Westmoreland, William. 1969. "Address to the Association of the US Army," October 14.

Wiener, Norbert. 1967. *The Human Use of Human Beings: Cybernetics and Society*. New York: Avon Books.

Wiggers, Kyle. 2021. "Continual Learning Offers a Path toward More Human-like AI." *Venture Beat*, April 9.

Wittgenstein, Ludwig. 1973. *Philosophical Investigations*. 3rd ed., translated by G. E. M. Anscombe. New York: Prentice-Hall.

Yudkowsky, Eliezer. 2008. "Artificial Intelligence as a Positive and Negative Factor in Global Risk." In *Global Catastrophic Risks*, edited by Nick Bostrom, and Milan M. Ćirković, 308–345. New York: Oxford University Press.